

# Molecular Anharmonicity: A Computer-Aided Treatment

R. G. DELLA VALLE,<sup>1</sup> L. HALONEN,<sup>2</sup> E. VENUTI<sup>1,2</sup>

<sup>1</sup>*Dipartimento di Chimica Fisica e Inorganica, Università di Bologna, Viale Risorgimento 4, I-40136 Bologna, Italy*

<sup>2</sup>*Laboratory of Physical Chemistry, P.O. Box 55 (A. I. Virtasen aukio 1), FIN-00014 University of Helsinki, Finland*

*Received 1 December 1998; accepted 21 July 1999*

**ABSTRACT:** We present a new software package for the theoretical treatment of anharmonic vibrational spectra of nonlinear polyatomic molecules. The package, called "B&D," computes vibrational energies starting from sets of force constants defined as potential energy derivatives. The method employed allows us to combine experimental rotation-vibration data with any information made available from *ab initio* calculations. The package follows the natural procedure in which a molecular problem is solved, both in the symbolic construction of Hamiltonian operator and basis functions and in the numerical computation of the Hamiltonian matrix elements. The novelty consists in making the entire procedure fully automatic, so that the occurrence of errors is greatly reduced and the laborious process involved in deriving and implementing the Hamiltonian is dramatically simplified. © 1999 John Wiley & Sons, Inc. *J Comput Chem* 20: 1716–1730, 1999

**Keywords:** molecular spectroscopy; anharmonic vibrations; group theory; Clebsch–Gordan coefficients; symbolic algebra

## Introduction

Theoretical methods of increasing complexity are needed to analyze and interpret in detail vibrational spectra of polyatomic molecules. Certainly, methods that give a description of the mole-

cular potential energy surface, retain, among the others, a formidable physical meaning. We follow here the approach of Halonen's group,<sup>1–4</sup> which includes an expansion of the Hamiltonian in a power series of the internal coordinates, its transformation to symmetry coordinates, the evaluation of the Hamiltonian in a matrix form with a basis of harmonic oscillator wavefunctions in symmetry coordinates, the transformation of this matrix to a symmetry adapted basis, and the diagonalization of the transformed matrix. This work describes a new software package developed to perform automati-

Correspondence to: R. G. Della Valle; e-mail: valle@aronte.fci.unibo.it

Contract/grant sponsors: The Academy of Finland, the Rector of the University of Helsinki, MURST, CNR, and the University of Bologna

cally all the steps required by the solution of the vibrational problem.

The package is called “build & deploy” (B&D), because its overall organization mirrors the natural order in which a new molecular problem is solved, and is decomposed in a “build” phase and in a “deploy” phase. The “build” program replaces the work that is usually done by hand (often with the aid of symbolic algebra programs) and, starting from a compact specification of the problem, writes an appropriate Fortran source. This source, combined with a fitting code and other auxiliary procedures, constitutes the “deploy” program, which performs the actual calculation and the fit of the spectra.

We have found that the rather unconventional strategy of writing a program that builds another program can be quite effective. In fact, the “build” program, once written, immediately works for any order of the anharmonic Hamiltonian, and for higher and higher overtones and combination states and, with a minimal effort, can be repeatedly applied to molecules of varying complexity. For methane isotopomers with  $T_d$  symmetry<sup>4</sup> we have automatically obtained  $\approx 3000$  terms up to the quartic Hamiltonian, and  $\approx 280$  states vectors in the symmetry adapted basis up to three stretchings or six bendings. We have fitted 80 experimental wave numbers, simultaneously on  $\text{CH}_4$ ,  $^{13}\text{CH}_4$ ,  $\text{CD}_4$ , and  $^{13}\text{CD}_4$  isotopomers, and obtained an R.M.S. deviation  $\sigma = 1.58 \text{ cm}^{-1}$ .

This article, which describes the theoretical and algorithmical aspects of the work, has been written simultaneously with the package, and follows the same top-down order with which the program has been designed. The theory and the top-level abstract algorithms appear first, followed by lower level algorithms and implementation details. The anharmonic Hamiltonian in internal coordinates and its transformations to symmetry coordinates and to second quantization operators are presented in the next section. The construction of the basis, the evaluation of the Hamiltonian in a matrix form and its transformations to a symmetry adapted basis are described in the Theory—The Basis Section. The specification of the molecular system is discussed in the Data Structures Section, together with the main features of the data structures.

The sequence of steps in the actual execution of the package, which differs from the order in which the methods are described, is summarized in Tables II and III in the Data Structures Section. The tables provide a reference frame for the following discussion, and it may be helpful to give them a

glance before reading the rest of the article. The B&D User’s Manual and examples for  $\text{NH}_3$  and  $\text{CH}_4$  molecules, which may also clarify the discussion, are available via the www.<sup>5</sup> A partial, but significant, example is also given in the Appendix.

Complications strictly related to a specific problem, such as the treatment of redundant coordinates, have been neglected in this article. These problems have been handled in a recent work,<sup>4</sup> where we have used the package to fit the overtone and combination spectra of the  $\text{CH}_4$  molecule.

## Theory—The Hamiltonian

### HAMILTONIAN IN INTERNAL COORDINATES

The treatment of harmonic and anharmonic vibrations of isolated molecules in the ground electronic state is discussed in many standard texts,<sup>3,6,7</sup> and will be briefly summarized here, to identify the problem and to introduce a nonstandard but useful terminology. The dynamics of a molecule performing small-amplitude oscillations about an equilibrium structure is best described in terms of curvilinear internal displacement valence coordinates,<sup>8</sup> which are nonlinear functions of the cartesian atomic coordinates. When the kinetic energy  $T$  and the potential energy  $V$  of the molecule are expressed in terms of the internal coordinates  $R_r$  and of their conjugate momenta  $P_r = -i\hbar\partial/\partial R_r$ , the vibrational ( $J = 0$ ) Hamiltonian reduces to:<sup>9,10</sup>

$$H = \frac{1}{2} \sum_{r_1 r_2} P_{r_1} G_{r_1 r_2}(\mathbf{R}) P_{r_2} + V(\mathbf{R}) \quad (1)$$

The Wilson’s  $G(\mathbf{R})$  matrix<sup>6</sup> and the Born–Oppenheimer<sup>11</sup> potential  $V(\mathbf{R})$  are functions of the internal coordinates. A small purely quantum mechanical kinetic energy term  $V'(\mathbf{R})$ , which is mass dependent but does not contain the momenta, is neglected.<sup>10</sup> The exact kinetic energy operator could be handled with numerical integration or discrete variable representation (DVR) methods,<sup>12,13</sup> while the potential, in most cases, is only known as a truncated power series in the coordinates. To simplify the calculations we expand both  $G(\mathbf{R})$  and  $V(\mathbf{R})$  in a power series of the displacement coordinates  $R_r$ , which allow us to treat even moderately large molecules, such as methane. The vibrational Hamiltonian  $H$  thus becomes:

$$H = H_2 + H_3 + H_4 + \cdots = \sum_{n \geq 2} H_n \quad (2)$$

$$H_n = T_n + V_n \quad (3)$$

$$T_n = \sum_{r_1 r_2 \dots r_{n-1} r_n} T_{r_1 r_2 \dots r_{n-1} r_n} P_{r_1} R_{r_2} \dots R_{r_{n-1}} P_{r_n}$$

$$V_n = \sum_{r_1 r_2 \dots r_{n-1} r_n} V_{r_1 r_2 \dots r_{n-1} r_n} R_{r_1} R_{r_2} \dots R_{r_{n-1}} R_{r_n} \quad (4)$$

Here  $T_{r_1 r_2 \dots r_{n-1} r_n}$  and  $V_{r_1 r_2 \dots r_{n-1} r_n}$  are constant coefficients whose indices  $r_j$ ,  $j = 1, 2, \dots, n$ , range over all the internal coordinates. The coefficients arise as derivatives of  $V(\mathbf{R})$  and  $G(\mathbf{R})$ , evaluated at the equilibrium structure ( $\mathbf{R} = 0$ ):

$$V_{r_1 r_2 \dots r_{n-1} r_n} = \frac{1}{n!} \left( \frac{\partial^n V(\mathbf{R})}{\partial R_{r_1} \dots \partial R_{r_n}} \right)_{\text{eq}}$$

$$T_{r_1 r_2 \dots r_{n-1} r_n} = \frac{1}{2} \frac{1}{(n-2)!} \left( \frac{\partial^{n-2} G_{r_1 r_n}(\mathbf{R})}{\partial R_{r_2} \dots \partial R_{r_{n-1}}} \right)_{\text{eq}} \quad (5)$$

All numeric divisors arising from the power series expansions have been incorporated directly into the definition of the coefficients. Thus, for example  $T_{r_1 r_2}$  and  $V_{r_1 r_2}$  are one half of Wilson's  $G_{r_1 r_2}$  and  $F_{r_1 r_2}$ , respectively.<sup>6</sup> This nonstandard choice enhances the mathematical symmetry of the Hamiltonian [eq. (4)], and avoids many divisors in following manipulations. We call  $T_{r_1 \dots r_n}$  a kinetic energy constant, and  $V_{r_1 \dots r_n}$  a potential energy constant, although the usual name for the latter is a force constant. When the distinction is unimportant, we refer to both of them as Hamiltonian constants.

Because the order of the derivatives is irrelevant, the potential energy constants  $V_{r_1 \dots r_n}$  are symmetric with respect to all  $n!$  permutations in the order of the labels  $r_1 \dots r_n$ . Each  $T_n$  term contains two momenta  $P_{r_1}$  and  $P_{r_n}$ , and  $n-2$  coordinates  $R_{r_2} \dots R_{r_{n-1}}$ . The terms with  $n > 2$  arise from the expansion of Wilson's  $G_{r_1 r_2}(\mathbf{R})$  matrix elements<sup>6</sup> in a power series of the displacements  $R_r$ , and are present solely because of the nonlinearity of the transformation between cartesian and internal coordinates. Because the  $G$  matrix is symmetric, and because of the properties of the derivatives, the kinetic energy constants  $T_{r_1 r_2 \dots r_{n-1} r_n}$  are symmetric for exchanges of the first and last labels  $r_1$  and  $r_n$  and for all  $(n-2)!$  permutations of the central labels  $r_2 \dots r_{n-1}$ .

### EQUIVALENT HAMILTONIAN CONSTANTS IN INTERNAL COORDINATES

Under the effect of the symmetry operators internal coordinates transform into linear combinations of internal coordinates. In the most general case, for example, for the bendings in a linear molecule, one coordinate goes into a combination of several coordinates. For stretchings and bendings of nonlinear molecules, however, each coordinate transforms into a single new coordinate, and a symmetry op-

eration is equivalent to a relabeling of the internal coordinates. The current version of the program is restricted to this case.

By exchanging internal coordinates, symmetry operators also exchange Hamiltonian constants. Exchanged constants must, of course, possess the same numerical value, being derivatives evaluated at the equilibrium structure [eq. (5)], by definition unaffected by the symmetry operations. When the molecule belongs to a nontrivial symmetry group the Hamiltonian  $H$  [eqs. (2)–(4)] can be simplified. For this purpose, we partition all constants into equivalence classes  $e$ , which are disjoint subsets containing constants with the same value.

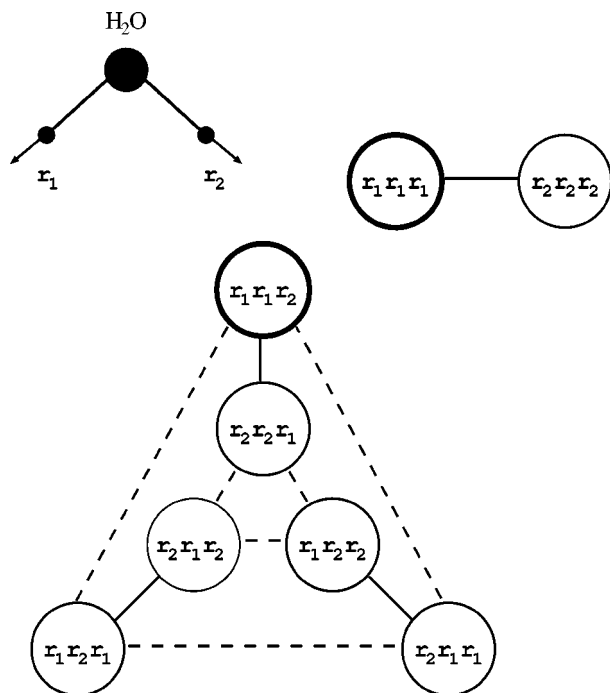
In simple cases, equivalent force constants can be identified with the help of geometric intuition. The program replaces human intuition with a graph-theoretical<sup>14,15</sup> algorithm, by representing Hamiltonian constants as vertices of a graph. Each constant  $V_{r_1 \dots r_n}$  or  $T_{r_1 \dots r_n}$ , and thus, the corresponding vertex, is identified by its  $n$  coordinate labels  $r_1 \dots r_n$ , as shown by the example in Figure 1. The permutation symmetry in the order of the labels, discussed in the previous section, and the molecular symmetry, translated into the relabeling of the internal coordinates, are systematically exploited to identify pairs of equivalent constants. All such pairs are recorded as directly connected pairs of graph vertices. Indirectly connected vertices, which are also equivalent because equivalence is transitive, are incrementally identified in this phase. The equivalence classes of Hamiltonian constants, which can be described as disjoint subsets of directly or indirectly connected vertices, are finally found by walking through all chains of connections.<sup>15</sup>

This algorithm, which at each order  $n$  has to be repeated separately for the potential and kinetic energy constants  $V_{r_1 \dots r_n}$  and  $T_{r_1 \dots r_n}$  because their permutation symmetries are different, results into a decomposition of the Hamiltonian constants into equivalence classes  $e$ , represented as linked lists of vertices.<sup>16</sup> The sums on the coordinates labels  $r_1 \dots r_n$  in  $H$  [eq. (4)] can thus be decomposed into sums over equivalence classes  $e$ :

$$T_n = \sum_e T_{ne} \sum_{r_1 r_2 \dots r_{n-1} r_n \in e} P_{r_1} R_{r_2} \dots R_{r_{n-1}} P_{r_n}$$

$$V_n = \sum_e V_{ne} \sum_{r_1 r_2 \dots r_{n-1} r_n \in e} R_{r_1} R_{r_2} \dots R_{r_{n-1}} R_{r_n} \quad (6)$$

where  $T_{ne}$  and  $V_{ne}$  are representative Hamiltonian constants of the class  $e$  of  $T_{r_1 \dots r_n}$  and  $V_{r_1 \dots r_n}$ . Because all the constants in a class possess the same value, any of them can be chosen to represent all the other



**FIGURE 1.** The H<sub>2</sub>O molecule (top, left) and the equivalence classes for the potential energy constants involving three stretching coordinates. Each graph vertex represents a  $V_{r_1 r_1 r_1}$  constant. Solid lines connect constants equivalent by virtue of the  $C_{2v}$  molecular symmetry. Broken lines connect constants equivalent because of the permutation symmetry of  $V_{r_1 r_1 r_1}$ . Directly and indirectly connected constants form disjoint equivalence classes. The first constant of each class, in lexicographic order, is chosen to be the canonical representative of the class itself (thick circles). In this example, two classes are found, one with the constants equivalent to  $V_{r_1 r_1 r_1}$  and the other with the constants equivalent to  $V_{r_1 r_1 r_2}$ . The corresponding Hamiltonian terms are  $V_{r_1 r_1 r_1}(r_1 r_1 r_1 + r_2 r_2 r_2)$  and  $V_{r_1 r_1 r_2} 3(r_1 r_1 r_2 + r_1 r_2 r_2)$ , respectively. The multiplicity factors (1 and 3, in this example) account for the number of terms equivalent by permutation symmetry.

ones. As explained in detail later, we systematically sort the constants in lexicographic order, and then choose the first of each class as the canonical representative. Examples of this decomposition of  $H$  appear in the caption of Figure 1, and in the Appendix.

### SYMMETRY COORDINATES AND PROJECTION OPERATORS

Following refs. 3, 6, and 7, the Hamiltonian is transformed from internal to symmetry coordinates  $\tilde{R}_s$ . These coordinates are linear combinations of the internal coordinates, chosen in such a way that un-

der the action of the symmetry operators each of them transform like a specific basis vector  $|li\rangle$  of an irreducible representation (IRREP)  $l$  of the group. The orientation label  $i$  distinguishes different basis vectors with the same  $l$ . Because  $H$  cannot be affected by any symmetry operator  $\hat{G}$ , only those few products of symmetry coordinates and momenta spanning the total-symmetric IRREP may be associated to nonzero Hamiltonian constants.

We build the symmetry coordinates by applying projection operators<sup>17, 18</sup> to internal coordinates. For this purpose we choose a specific set of matrices  $D^l(\hat{G})$  for the IRREPs  $l$  of the group. These matrices contain matrix elements of the symmetry operators  $\hat{G}$  between basis vectors  $|li\rangle$  and  $|lj\rangle$  of the IRREPs:

$$D^l_{ij}(\hat{G}) = \langle li | \hat{G} | lj \rangle \quad (7)$$

The choice of  $D^l(\hat{G})$  matrices is unique for one-dimensional IRREPs. For multidimensional IRREPs, one can choose different combinations of basis vectors, and thus obtain different matrices. Although all choices are equivalent, some care is required<sup>5</sup> to reproduce published sets of symmetry coordinates. To build the symmetry coordinates we use the projector operators  $\hat{P}^l_{ij}$ , which are linear combinations of the group operators, defined as

$$\hat{P}^l_{ij} = \sum_{\hat{G}} D^l_{ij}(\hat{G}) \hat{G} \quad (8)$$

Because we normalize the result of the projection, we have dropped some numeric factors that normally appear in the definition of  $\hat{P}^l_{ij}$ . There are two types of projectors:<sup>17</sup>  $\hat{P}^l_{ii}$ , which projects an arbitrary vector  $|r\rangle$  on  $|li\rangle$ ; and  $\hat{P}^l_{ij}$ , with  $i \neq j$ , which rotates  $|lj\rangle$  onto  $|li\rangle$ . Apparently one could use  $\hat{P}^l_{ii}$  to build a complete  $|li\rangle$  basis out of a set of initial vectors  $|r\rangle$ , but this is not the case. In fact, the projection is uniquely determined only up to a sign, or, for complex  $D^l(\hat{G})$  matrices, up to a phase. Thus two vectors  $|li\rangle = \hat{P}^l_{ii}|r\rangle$  and  $|lj\rangle = \hat{P}^l_{jj}|r\rangle$  do not necessarily satisfy eq. (7), because they might have inconsistent phases. To avoid this problem only the first basis vector of each IRREP  $l$ ,  $|l1\rangle = \hat{P}^l_{11}|r\rangle$  is built directly, while the other basis vectors  $|li\rangle$ , for multi-dimensional IRREPs, are obtained indirectly by rotating  $|l1\rangle$ :  $|li\rangle = \hat{P}^l_{i1}|l1\rangle$ . The relative phases of the vectors so obtained satisfy eq. (7) automatically.

To complete the specification of the symmetry coordinates, beside the projectors, one has to choose the initial vectors  $|r\rangle$ , or, in this context, the initial internal coordinates  $R_r$ . We use a blind machine-oriented method, and sequentially apply all projectors  $\hat{P}^l_{11}$  to all internal coordinates  $R_r$ , and all  $\hat{P}^l_{i1}$

to  $\hat{P}_{11}^l R_r$ . The result of each projection is orthonormalized with the Gram–Schmidt method<sup>19</sup> to all projections already obtained. Zero projections and projections that are linear combinations of previously computed projections are automatically eliminated by this procedure. The final result of the method is a square matrix  $J$  that transforms the internal coordinates  $R$  into symmetry coordinates  $\tilde{R} = JR$ :

$$\tilde{R}_s = \sum_r J_s^r R_r \quad (9)$$

The orthonormalization procedure guarantees the unitarity ( $J^{-1} = J^\dagger$ ) of the Jacobian matrix  $J$  (which has components  $J_s^r = \partial \tilde{R}_s / \partial R_r$ ), and thus allows one to express the internal coordinates  $R_r$  and associated momenta  $P_r$  as a linear combination of symmetry coordinates  $\tilde{R}_s$  and momenta  $\tilde{P}_s$ :

$$R_r = \sum_s J_s^r \tilde{R}_s \quad (10)$$

$$P_r = \sum_s J_s^r \tilde{P}_s \quad (11)$$

Equations (9)–(11) need to be modified when redundancies occur, i.e., when the number of internal coordinates is larger than that of internal degrees of freedom of the molecule.<sup>6,7</sup> In these cases eq. (10) is to be replaced by a nonlinear relation,<sup>3,4,8,20</sup> whereas eqs. (9) and (11) maintain their linear structure, though in terms of a reduced number of coordinates  $\tilde{R}_s$  and momenta  $\tilde{P}_s$ .

## HAMILTONIAN IN SYMMETRY COORDINATES

The transformation of  $H$  from internal to symmetry coordinates is particularly simple when there are no redundancies and the original coordinates and momenta  $R_r$  and  $P_r$  are linear combinations of the new coordinates and momenta  $\tilde{R}_s$  and  $\tilde{P}_s$ , i.e., when eqs. (10) and (11) hold. By direct substitution of  $R_r$  and  $P_r$  in eq. (4), one obtains:

$$T_n = \sum_{s_1 s_2 \dots s_{n-1} s_n} \tilde{T}_{s_1 s_2 \dots s_{n-1} s_n} \tilde{P}_{s_1} \tilde{R}_{s_2} \dots \tilde{R}_{s_{n-1}} \tilde{P}_{s_n}$$

$$V_n = \sum_{s_1 s_2 \dots s_{n-1} s_n} \tilde{V}_{s_1 s_2 \dots s_{n-1} s_n} \tilde{R}_{s_1} \tilde{R}_{s_2} \dots \tilde{R}_{s_{n-1}} \tilde{R}_{s_n} \quad (12)$$

$$\tilde{T}_{s_1 s_2 \dots s_n} = \sum_{r_1 r_2 \dots r_n} T_{r_1 r_2 \dots r_n} J_{s_1}^{r_1} J_{s_2}^{r_2} \dots J_{s_n}^{r_n}$$

$$\tilde{V}_{s_1 s_2 \dots s_n} = \sum_{r_1 r_2 \dots r_n} V_{r_1 r_2 \dots r_n} J_{s_1}^{r_1} J_{s_2}^{r_2} \dots J_{s_n}^{r_n} \quad (13)$$

The equations for  $T_n$  and  $V_n$  have the same structure as those in eq. (4), with new constants  $\tilde{T}_{s_1 s_2 \dots s_n}$

and  $\tilde{V}_{s_1 s_2 \dots s_n}$  instead of  $T_{r_1 r_2 \dots r_n}$  and  $V_{r_1 r_2 \dots r_n}$ . Equation (13) represents a linear relationship between Hamiltonian constants in symmetry and internal coordinates, and is strictly valid only in the absence of redundancies; in fact, additional contributions to  $\tilde{T}_{s_1 s_2 \dots s_n}$  and  $\tilde{V}_{s_1 s_2 \dots s_n}$  are needed<sup>3,4</sup> when redundancies are present and eq. (10) does not hold. We report in ref. 4 the treatment for the redundancy<sup>6,7</sup> of the bendings in  $\text{CH}_4$ , together with the analytical expressions for the corrected kinetic energy quartic coefficients  $\tilde{T}_{s_1 s_2 s_3 s_4}$ .

Because the sums over Hamiltonian constants can be decomposed into sums over their equivalence classes  $e$  [eq. (6)], we repeat the same decomposition in eq. (13), obtaining:

$$\tilde{T}_{s_1 s_2 \dots s_n} = \sum_e T_{ne} X_{s_1 s_2 \dots s_n}^e$$

$$\tilde{V}_{s_1 s_2 \dots s_n} = \sum_e V_{ne} X_{s_1 s_2 \dots s_n}^e \quad (14)$$

$$X_{s_1 s_2 \dots s_n}^e = \sum_{r_1 r_2 \dots r_{n-1} r_n \in e} J_{s_1}^{r_1} J_{s_2}^{r_2} \dots J_{s_n}^{r_n} \quad (15)$$

Equations (14) and (15) describe in a clear way a complete separation between the geometric and nongeometric aspects of  $H$ . The dimensionless coefficients  $X_{s_1 s_2 \dots s_n}^e$  have a purely geometric origin and depend only on the molecular symmetry. All nongeometric properties of the interaction are concentrated in the strength of the couplings  $T_{ne}$  and  $V_{ne}$ . The normal way of expressing the constants in symmetry coordinates as a function of the distinct constants in internal coordinates [eq. (14)] is to replace  $R_r$  and  $P_r$  [eqs. (10) and (11)] in  $H$  [eq. (6)] and then to rearrange term by term the resulting expression to obtain the desired explicit expression. The process is laborious, even with the assistance of a symbolic algebra program.

The expression for the transformation coefficients in eq. (15) is already in a completely explicit form, and is suitable for a purely numeric calculation, which is feasible at any Hamiltonian order. The inputs to the calculation are the Jacobian matrix  $J$ , which connects symmetry and internal coordinates, and the equivalence classes of the Hamiltonian constants in internal coordinates, represented as lists of coordinate indices  $r_1 \dots r_n$ . The output is the transformation of the constants from internal to symmetry coordinates. The program also recognizes in this phase those Hamiltonian constants that are multiples or linear combinations of other constants, so that the final Hamiltonian operator of eq. (14) is expressed in terms of linearly independent constants in symmetry coordinates. The number and type of

independent nonvanishing Hamiltonian constants thus obtained have been found to coincide with those given by other methods.<sup>21, 22</sup>

## TERM MULTIPLICITY

The number of terms in the Hamiltonian [eq. (12)] may be reduced by exploiting the permutation symmetry of the terms. The basic strategy is: given a set of  $m$  terms equivalent by permutation symmetry, keep only a single canonical representative term and multiply it by the multiplicity  $m$ . Therefore eq. (12) remains unaltered, except that the coefficients  $\tilde{T}_{s_1 s_2 \dots s_{n-1} s_n}$  and  $\tilde{V}_{s_1 s_2 \dots s_{n-1} s_n}$  incorporate multiplicity factors and the sums on  $s_1 s_2 \dots s_{n-1} s_n$  are restricted to representative terms. To implement the strategy, we need an arbitrary criteria for choosing a representative term, and a method for counting the number of equivalent permutations. As a representative term of a set we choose the first one in lexicographic order, as we did for the Hamiltonian constants shown above. A detailed description will be presented below.

To find the number of equivalent terms, we have to treat separately the kinetic and potential terms, because their permutation symmetries are different. Potential terms,  $\tilde{V}_{s_1 \dots s_n} \tilde{R}_{s_1} \dots \tilde{R}_{s_n}$ , are particularly simple, because both the  $\tilde{V}_{s_1 \dots s_n}$  constants and the operator products  $\tilde{R}_{s_1} \dots \tilde{R}_{s_n}$  are invariant under all permutations of the  $n$  coordinate labels  $s_1 \dots s_n$ . To find the number of distinct permutations without actually generating and counting them directly, we use the method of multisets described by Knuth.<sup>23</sup> A multiset is like a set, except that it can have repetitions of identical elements. If all  $n$  labels are different, the number of permutations is  $n!$ . For every subset of  $n_i$  identical labels, one has to divide by  $n_i!$  to find the number of distinct permutations.

Because kinetic terms,  $\tilde{T}_{s_1 s_2 \dots s_{n-1} s_n} \tilde{P}_{s_1} \tilde{R}_{s_2} \dots \tilde{R}_{s_{n-1}} \tilde{P}_{s_n}$ , are symmetric for permutations of the coordinate labels  $s_2 \dots s_{n-1}$ , their multiplicity for such permutations is computed like the multiplicity of the potential energy terms. The multiplicity arising from the momenta  $\tilde{P}_{s_1}$  and  $\tilde{P}_{s_n}$ , is complicated by a possible mismatch between the symmetry of the constants and that of the corresponding operators. For example,  $\tilde{T}_{112} = \tilde{T}_{211}$ , but  $\tilde{P}_1 \tilde{R}_1 \tilde{P}_2 \neq \tilde{P}_2 \tilde{R}_1 \tilde{P}_1$ , because  $\tilde{P}_1$  and  $\tilde{R}_1$  do not commute. The exchange of  $\tilde{P}_{s_1}$  and  $\tilde{P}_{s_n}$  in an  $n$ th order kinetic term is allowed only if the momenta labels  $s_1$  and  $s_n$  never appear among the coordinate labels  $s_2 \dots s_{n-1}$ . If  $s_1 \neq s_n$  and the exchange is allowed, we keep only a single permutation, with multiplicity 2, otherwise we keep both permutations, with multiplicity 1. If  $s_1 = s_n$  the

exchange is pointless, and the multiplicity arising from the momenta is 1.

A formal definition of the canonical order for the term labels is now possible. For potential terms,  $\tilde{V}_{s_1 \dots s_n} \tilde{R}_{s_1} \dots \tilde{R}_{s_n}$ , we sort the coordinate labels  $s_1 \dots s_n$  in increasing numeric sequence. For kinetic terms,  $\tilde{T}_{s_1 s_2 \dots s_{n-1} s_n} \tilde{P}_{s_1} \tilde{R}_{s_2} \dots \tilde{R}_{s_{n-1}} \tilde{P}_{s_n}$ , we always sort  $s_2 \dots s_{n-1}$ . If  $s_1 \neq s_n$  and the exchange is allowed, we also sort  $s_1$  and  $s_n$ .

It should be noticed that the above discussion could also apply to the Hamiltonian in internal coordinates, eq. (6). However in eqs. (13) and (15) we need to sum explicitly on all equivalent permutations of internal coordinates. Therefore, we do not find convenient to replace equivalent permutations with multiplicity factors while working with  $H$  in internal coordinates; we do, however, perform the replacement while printing out the symbolic representation of  $H$ . An example of this procedure appears in the caption of Figure 1.

## CREATION AND ANNIHILATION OPERATORS

The products of  $\tilde{R}_s$  and  $\tilde{P}_s$  operators appearing in the Hamiltonian [eq. (12)] are most efficiently manipulated by expressing them in terms of harmonic oscillator annihilation and creation operators  $a_s^-$  and  $a_s^+$ . This transformation<sup>11, 24</sup> is defined by

$$\begin{aligned}\tilde{R}_s &= \alpha_s^{-1/2} 2^{-1/2} (a_s^+ + a_s^-) = \alpha_s^{-1/2} 2^{-1/2} \sum_{\sigma=\pm} a_s^\sigma \\ \tilde{P}_s &= i\hbar \alpha_s^{1/2} 2^{-1/2} (a_s^+ - a_s^-) \\ &= i\hbar \alpha_s^{1/2} 2^{-1/2} \sum_{\sigma=\pm} \sigma a_s^\sigma\end{aligned}\quad (16)$$

Here,  $\alpha_s$  is a free parameter that will be eventually fixed. The sums on positive and negative signs  $\sigma$  in eq. (16) are a compact notation introduced for later convenience. From eq. (16) it can be easily proven<sup>11, 24</sup> that the eigenvalues of the number operator  $\hat{n}_s = a_s^+ a_s^-$  are nonnegative integers  $n_s$ , and that the effect of  $a_s^+$  and  $a_s^-$  on the eigenstates  $|n_s\rangle$  of  $\hat{n}_s$  is:

$$\begin{aligned}a_s^+ |n_s\rangle &= (n_s + 1)^{1/2} |n_s + 1\rangle \\ a_s^- |n_s\rangle &= n_s^{1/2} |n_s - 1\rangle\end{aligned}\quad (17)$$

The states  $|n_s\rangle$  can also be defined, more physically, as the eigenfunctions  $\psi_{n_s}(\tilde{R}_s)$  of an harmonic oscillator Hamiltonian in symmetry coordinates

$$H_{ss} = \frac{1}{2} (\tilde{G}_{ss} \tilde{P}_s^2 + \tilde{F}_{ss} \tilde{R}_s^2) \quad (18)$$

This is just the  $s_1 = s_2 = s$  term of  $H_2 = T_2 + V_2$  in eq. (12), with a Wilson's matrix element  $\tilde{G}_{ss}$  and by a force constant  $\tilde{F}_{ss}$ . The corresponding harmonic

frequency is  $\omega_s = (\tilde{F}_{ss}\tilde{G}_{ss})^{1/2}$ , which corresponds to a wave number  $\hbar\omega_s/\hbar c_0$ . Here  $c_0$  is the speed of light in vacuum and  $\hbar$  is Planck constant. If eq. (16) is substituted in (18), one finds terms involving  $a_s^+a_s^-$ ,  $a_s^-a_s^+$ ,  $a_s^+a_s^+$  and  $a_s^-a_s^-$ . These latter two terms, which are not diagonal in the basis  $|n_s\rangle$ , cancel out if  $\alpha_s$  is fixed as:<sup>1,25</sup>

$$\alpha_s = \frac{1}{\hbar}(\tilde{F}_{ss}/\tilde{G}_{ss})^{1/2} \quad (19)$$

This standard choice makes the Hamiltonian  $H_{ss}$  diagonal in  $|n_s\rangle = \psi_{n_s}(\tilde{R}_s)$ , with energy eigenvalues  $\hbar\omega_s(n_s + \frac{1}{2})$ .

## SPECTROSCOPIC CONSTANTS

The matrix elements of  $H$  will be evaluated between harmonic oscillator wave functions  $|n_s\rangle$ , and this makes the operators  $a_s^\pm$  more convenient than  $\tilde{R}_s, \tilde{P}_s$ . Therefore we substitute eq. (16) into (12), and obtain

$$T_n = \sum_{\substack{\sigma_1\sigma_2\ldots\sigma_{n-1}\sigma_n \\ s_1s_2\ldots s_{n-1}s_n}} \Gamma_{\sigma_1\sigma_2\ldots\sigma_{n-1}\sigma_n}^{\sigma_1\sigma_2\ldots\sigma_{n-1}\sigma_n} a_{s_1}^{\sigma_1} a_{s_2}^{\sigma_2} \ldots a_{s_{n-1}}^{\sigma_{n-1}} a_{s_n}^{\sigma_n} \\ V_n = \sum_{\substack{\sigma_1\sigma_2\ldots\sigma_{n-1}\sigma_n \\ s_1s_2\ldots s_{n-1}s_n}} \Phi_{\sigma_1\sigma_2\ldots\sigma_{n-1}\sigma_n}^{\sigma_1\sigma_2\ldots\sigma_{n-1}\sigma_n} a_{s_1}^{\sigma_1} a_{s_2}^{\sigma_2} \ldots a_{s_{n-1}}^{\sigma_{n-1}} a_{s_n}^{\sigma_n} \quad (20)$$

$$\Gamma_{\sigma_1\sigma_2\ldots\sigma_{n-1}\sigma_n}^{\sigma_1\sigma_2\ldots\sigma_{n-1}\sigma_n} = -\hbar^2 \sigma_1 \sigma_n 2^{-n/2} \alpha_{s_1}^{1/2} \alpha_{s_2}^{-1/2} \ldots \\ \times \alpha_{s_{n-1}}^{-1/2} \alpha_{s_n}^{1/2} T_{s_1s_2\ldots s_{n-1}s_n} \\ \Phi_{\sigma_1\sigma_2\ldots\sigma_{n-1}\sigma_n}^{\sigma_1\sigma_2\ldots\sigma_{n-1}\sigma_n} = 2^{-n/2} \alpha_{s_1}^{-1/2} \alpha_{s_2}^{-1/2} \ldots \\ \times \alpha_{s_{n-1}}^{-1/2} \alpha_{s_n}^{-1/2} V_{s_1s_2\ldots s_{n-1}s_n} \quad (21)$$

The matrix elements of  $T_n$  and  $V_n$  [eq. (20)] between  $|n_s\rangle$  wave functions are particularly simple, because only the dependence on the quantum numbers  $n_s$  remains in the dimensionless operators  $a_s^\pm$ . All other quantities have been incorporated into the coefficients  $\Gamma$  and  $\Phi$ , which have the dimension of an energy and are related to the spectroscopic constants. The sums on all sign pairs,  $\sigma_i = \pm$ , constitute a bookkeeping device that keeps track of all  $2^n$  combinations of  $a_{s_i}^+, a_{s_i}^-$  operators in each  $T_n, V_n$  operator. This device, originally developed for solid state applications,<sup>26</sup> has been successfully used to handle anharmonic perturbations up to  $V_7$  in solid nitrogen.<sup>27</sup>

## Theory—The Basis

### CHOICE OF THE BASIS

If the series in eq. (2) is truncated at  $n = 2$ , the Hamiltonian can be diagonalized exactly by using

the familiar Wilson's *GF* method,<sup>6</sup> which yields a set of normal modes. Exact diagonalization is no more possible for  $n > 2$ . One is forced to evaluate the Hamiltonian in a matrix form, after choosing a suitable basis  $B$  of quantum states, and to diagonalize the matrix numerically. Products  $|n_1 n_2 \ldots\rangle = \psi_{n_1}(\tilde{R}_1) \psi_{n_2}(\tilde{R}_2) \ldots$  of harmonic oscillator eigenfunctions  $|n_s\rangle = \psi_{n_s}(\tilde{R}_s)$  in symmetry coordinates  $\tilde{R}_s$  are a reasonable choice as basis states, because they transform in particularly simple ways [eq. (17)] under the action of the  $a_s^\pm$  operators contained in  $H$  [eq. (20)].

Because infinite basis sets, which would yield exact results, cannot be handled in practice, one has to divide the basis states in blocks of finite size. Subsets of "resonant" states close in energy, which mix strongly, are the best choice for the states forming a block, in the sense that this choice minimizes the numerical error. The restriction to resonant blocks results in a partition of the Hamiltonian  $H$  into intrablock terms  $H^0$  and interblock terms  $H^1$ , with  $H = H^0 + H^1$ . Quasiresonant terms, which couple pairs of states within a single block, are in  $H^0$ , while all other terms, which necessarily couple pairs of states belonging to different blocks, are in  $H^1$ .

A common example of this situation ( $\text{H}_2\text{O}$ ,  $\text{NH}_3$ ,  $\text{CH}_4$ , etc.) is the Fermi resonance between X—H stretching ( $r$ ) and H—X—H bending ( $\theta$ ) frequencies,  $\omega_r \approx 2\omega_\theta$ . The states with  $2N_r + N_\theta = \text{constant}$ , where  $N_r$  and  $N_\theta$  are the total number of  $r$  and  $\theta$  quanta, form resonant subsets of states and are a sensible choice as basis blocks.

## HAMILTONIAN IN A MATRIX FORM

After choosing a constraint for the quantum numbers  $n_s$  in a basis block, all the states  $|n_1 n_2 \ldots\rangle$  satisfying the constraint are generated by solving a simple combinatorial problem. The matrix elements  $H_{b'b} = \langle b'|H|b\rangle$  of  $H$  [eq. (20)] between all pairs  $b', b$  of states in the block are then found by applying each term of  $H$  to each state in the basis  $B$ . The action of a term is easily determined by applying all operators  $a_{s_i}^{\sigma_i}$  in the term, according to eq. (17), and yields a final state  $|b'\rangle = |n'_1 n'_2 \ldots\rangle$ , multiplied by a numeric coefficient. If the final state is still in the initial basis block, the coefficient is added to the matrix element  $H_{b'b}$ , otherwise is discarded. When all  $H$  terms and all states in the block have been exhausted, we are left with a square symmetric array  $H$ , which represents  $H$  in a matrix form.

## SECOND-ORDER PERTURBATIONS

A minor extension of the previous algorithm is sufficient to handle second-order perturbations. Be-

cause the basis blocks are made of states close in energy, and  $H$  has been partitioned into intrablock and interblock terms,  $H = H^0 + H^1$ , the standard Van Vleck<sup>11</sup> perturbation formula for the coupling between nearly degenerate states  $b'$  and  $b$  reduces to the effective matrix element

$$H_{b'b} = \langle b'|H^0|b\rangle + \frac{1}{2} \sum_{b''} \langle b'|H^1|b''\rangle \langle b''|H^1|b\rangle \times \left( \frac{1}{E_{b'}^0 - E_{b''}^0} + \frac{1}{E_b^0 - E_{b''}^0} \right) \quad (22)$$

where  $E_b^0 = \langle b|H^0|b\rangle$  is the unperturbed energy of the state  $b$ . Since  $H^1$  contains only interblock terms,  $b''$  does not belong to the same block of  $b$  and  $b'$ . In these conditions the perturbation theory is generally applicable, because the resonance denominators,  $|E_{b'}^0 - E_{b''}^0|$  and  $|E_b^0 - E_{b''}^0|$ , are necessarily much larger than  $|E_{b'}^0 - E_b^0|$ . The calculation of the second-order contribution to  $H_{b'b}$  is like that of the first-order contribution described in the previous section, except that the Hamiltonian  $H^1$  is applied twice and the energy denominators must be included.

It is worth noticing that, although we can expand the Hamiltonian to arbitrary orders as a power series of the coordinates, we are, in the present implementation, limited to second-order Van Vleck perturbation theory<sup>11</sup> when treating interblock interactions. An extension of the program, for example, using the algebraic approach of Sibert and McCoy,<sup>28,29</sup> would allow arbitrary perturbation orders to be taken into account.

### SYMMETRY-ADAPTED BASIS AND CLEBSH-GORDAN COEFFICIENTS

By dividing the basis states into near resonant blocks, and by treating interblock interactions as perturbations, an effective block diagonal Hamiltonian matrix  $H$  is obtained. The matrix  $H$ , which has been evaluated in a basis  $B$  of products  $|n_1 n_2 \dots\rangle$  of  $\psi_n(\tilde{R}_s)$  wave functions in symmetry coordinates  $\tilde{R}_s$ , has to be diagonalized numerically for each basis block. A significant reduction of the numeric effort, and some theoretical advantage, can be gained if  $H$ , which is already factorized by resonant basis blocks, is further factorized to even finer symmetry blocks, by transforming it to a symmetry adapted basis  $C$ . The states in the symmetry adapted basis  $C$  are linear combinations of the  $B$  states, chosen to transform as basis vectors of the group IRREPs under the action of the symmetry operators.

We build the symmetry adapted states in the same way as the symmetry coordinates, by applying projection operators to the  $|n_1 n_2 \dots\rangle$  states. The

$|n_1 n_2 \dots\rangle$  states can be obtained by repeatedly applying  $a_s^+$  creation operators (which transform in the same way as the corresponding symmetry coordinate  $\tilde{R}_s$ <sup>17,24</sup>) to the vibrational ground state  $|00 \dots 0\rangle$ , which is totally symmetric by definition. This is enough to define the action of group operator  $\hat{G}$  on a single  $|n_1 n_2 \dots\rangle$  state.

Because the numeric label  $s$  of the symmetry coordinate  $\tilde{R}_s$  does not contain the IRREP information, in this section we switch to a more explicit notation. We denote with  $a_{fj}^+$  a creation operator that transforms as the  $j$ th orientation of the  $l$ th IRREP. The family index  $f$  distinguishes between different operators with same  $l$  and  $j$  (i.e., between different coordinates belonging to the same symmetry species). We indicate the ground state with  $| \rangle$ , the first vibrational state  $a_{fj}^+ | \rangle$  with  $|fj\rangle$ , and the combination state  $a_{f_1 l_1 j_1}^+ \dots a_{f_n l_n j_n}^+ | \rangle$  with  $|f_1 l_1 j_1 \dots f_n l_n j_n\rangle$ . The number of  $fj$  triplets in the ket is thus the total number of excited vibrational quanta. For example a fundamental stretching  $r$  of  $E$  symmetry,  $x$  component, will be represented as  $|rE_x\rangle$ , while  $|rE_x rE_y\rangle$  will denote a combination state.

The notations  $|fj \dots\rangle$  and  $|n_1 n_2 \dots\rangle$  are simply alternative ways of naming the same state. The former notation is convenient when symmetry is an issue, the latter is more suitable to evaluate matrix elements of the Hamiltonian. In the  $|fj \dots\rangle$  notation it is easy to describe the effect of  $\hat{G}$  and  $\hat{P}_{ij}^l$  operators. Because  $\hat{G}$  can only couple components within the same IRREP  $l$  and family  $f$ , eq. (7) translates into

$$\hat{G}|fj\rangle = \sum_i |fli\rangle D_{ij}^l(\hat{G}) \quad (23)$$

as one can verify by multiplying with  $\langle fli|$ . From eq. (23) it follows immediately

$$\hat{G}|f_1 l_1 j_1 \dots f_n l_n j_n\rangle = \sum_{i_1 \dots i_n} |f_1 l_1 i_1 \dots f_n l_n i_n\rangle D_{i_1 j_1}^{l_1}(\hat{G}) \dots D_{i_n j_n}^{l_n}(\hat{G}) \quad (24)$$

Let us indicate the projection of  $|f_1 l_1 j_1 \dots f_n l_n j_n\rangle$  onto the  $|li\rangle$  subspace as

$$|li/f_1 l_1 j_1 \dots f_n l_n j_n\rangle = \hat{P}_{ii}^l |f_1 l_1 j_1 \dots f_n l_n j_n\rangle \quad (25)$$

where the projector  $\hat{P}_{ii}^l = \sum_{\hat{G}} D_{ii}^l(\hat{G}) \hat{G}$  has been used [eq. (8)]. If eqs. (8) and (24) are substituted in eq. (25), one obtains:

$$|li/f_1 l_1 j_1 \dots f_n l_n j_n\rangle = \sum_{i_1 \dots i_n} |f_1 l_1 i_1 \dots f_n l_n i_n\rangle \langle l i_1 \dots l i_n | li/f_1 \dots j_n \rangle \quad (26)$$



$$\langle l_1 i_1 \dots l_n i_n | l_i / j_1 \dots j_n \rangle = \sum_{\hat{G}} D_{ii}^l(\hat{G}) D_{i_1 j_1}^{l_1}(\hat{G}) \dots D_{i_n j_n}^{l_n}(\hat{G}) \quad (27)$$

The symbol  $\langle l_1 i_1 \dots l_n i_n | l_i / j_1 \dots j_n \rangle$  is just an abbreviation for the scalar product  $\langle f_1 l_1 i_1 \dots f_n l_n i_n | l_i / j_1 j_1 \dots f_n l_n j_n \rangle$ , which is acceptable if one realizes that the product is independent of the family indices  $f$  [eq. (27)] and that it is unnecessary to carry the IRREP indices  $l$  on both sides of the symbol. The coefficients  $\langle l_1 i_1 \dots l_n i_n | l_i / j_1 \dots j_n \rangle$  are called Clebsh–Gordan coefficients, Wigner coefficients, or vector addition coefficients.<sup>11, 17, 18, 24</sup>

Equations (26) and (27) describe the projection onto  $|l_i\rangle$  of a single product  $|f_1 l_1 j_1 \dots f_n l_n j_n\rangle$  of wave functions. In the presence of degenerate wave functions eq. (27) is not sufficient to obtain a symmetry adapted basis vector. In fact, if two wave functions are degenerate by symmetry, i.e., if they belong to the same family  $f$  and IRREP  $l$ , but have different orientations, say  $j_1$  and  $j_2$ , one has to account for the fact that  $|f l j_1 f l j_2\rangle$  and  $|f l j_2 f l j_1\rangle$  are actually the same state, because the permutation of the orientations among two or more degenerate wave functions does not change a product state. By simply summing together all such permutations with the proper normalization, we compute in a very direct and efficient way the symmetrized direct product.<sup>18</sup>

To build the complete symmetry adapted basis  $C$ , we repeat the projection and symmetrization procedure for all vectors of the original basis, and systematically orthonormalize each new symmetry adapted vector with respect to all vectors already obtained. The final result of the algorithm is a square unitary matrix  $C$  of Clebsh–Gordan coefficients, which connect the old and the new basis. An example can be found in Table I. In the new basis the Hamiltonian matrix  $H$  becomes  $C^\dagger H C$ , in the desired block diagonal form. Each block, which must be diagonalized numerically to find the vibrational wave numbers, corresponds to a single IRREP  $l$  and orientation  $i$ . All crossterms have been eliminated, and the computational effort is greatly reduced.

The problem can be simplified even further for molecules with noncommutative symmetry groups, which necessarily possess some degenerate symmetry species, i.e., IRREPs  $l$  with dimensionality larger than one. Each orientation  $i$  within such an IRREP exhibits the same set of wave numbers, so that a single  $i$  block, say  $i = 1$ , can be selected and all the others discarded. Practically, the symmetry adapted basis  $C$  is pruned by keeping only the subspace with  $i = 1$ . All vectors of the original basis  $B$  with a zero projection on the  $i = 1$  subspace can also be dis-

carded, because elements of the  $H$  matrix involving such  $B$  vectors contribute only to  $C$  subspaces with  $i \neq 1$ , which are degenerate with the  $i = 1$  subspace. An example of this procedure can be found in the caption of Table I.

Before leaving this section, we want to mention that the Clebsh–Gordan coefficients contain arbitrary phase factors<sup>11, 18</sup> whose choice is a matter of convention. Changes in the phases of the basis vectors do not affect the matrix eigenvalues, but may cause apparent differences with published coefficients.

## Data Structures and Algorithmical Details

### SYSTEM SPECIFICATION AND OBJECTS

The molecular system is fully described by complete sets of symmetry operators  $\hat{G}_g$ , of atoms  $A_a$  and of internal coordinates  $R_r$ . Here  $g$ ,  $a$ , and  $r$  are integer labels. For each operator  $\hat{G}_g$  in the symmetry group one needs the natural  $3 \times 3$  representation matrix  $D(\hat{G}_g)$  in a cartesian frame and all IRREP matrices  $D^l(\hat{G}_g)$ . The information on the complete group may be obtained from that of a properly chosen subset of operators (the group generators) by computing all the possible operator products. For each atom  $A_a$  in the system one needs the coordinates in a cartesian frame. The information on a properly chosen subset of atoms (the atom generators) is sufficient as the other atoms may be obtained by symmetry. A complete set of the internal coordinates  $R_r$ , for example, stretchings  $r_{ab}$ , bendings  $\theta_{abc}$ , etc., is needed. A proper subset of coordinates (the coordinate generators) is, again, sufficient.

The complete sets of group operators, of atomic coordinates, and of internal coordinates, are obtained from the corresponding generators by repeatedly applying the group operators, i.e., by computing the closure with respect to the group. We have found that the closure operation occupies a central role in the treatment of the problem. The definition of the equivalence classes, the block structure of the transformation from internal to symmetry coordinates, and even the high level specification of the data structures used by the program, all descend in a natural way from the notion of closure of a set of objects. We call an object any entity  $O_o$  to which a group operator  $\hat{G}_g$  can be applied yielding another entity  $O_{o'} = \hat{G}_g O_o$ . Atoms, internal coordinates, Hamiltonian constants, and group operators all qualify as objects.

**TABLE I.**  
**Basis States Involving Six Bendings of the CH<sub>4</sub> Molecule (*T<sub>d</sub>* Symmetry).**

Symmetry	$ \theta_{F_{2x}}^2 \theta_{F_{2y}}^4\rangle$  240)	$ \theta_{F_{2x}}^2 \theta_{F_{2z}}^4\rangle$  204)	$ \theta_{F_{2x}}^4 \theta_{F_{2y}}^2\rangle$  420)	$ \theta_{F_{2x}}^4 \theta_{F_{2z}}^2\rangle$  402)	$ \theta_{F_{2y}}^2 \theta_{F_{2z}}^4\rangle$  024)	$ \theta_{F_{2y}}^4 \theta_{F_{2z}}^2\rangle$  042)
A <sub>1</sub>	1/√6	1/√6	1/√6	1/√6	1/√6	1/√6
A <sub>2</sub>	1/√6	−1/√6	−1/√6	1/√6	1/√6	−1/√6
E <sub>x</sub>	.	1/2	.	−1/2	1/2	−1/2
E <sub>x</sub>	2/√12	−1/√12	2/√12	−1/√12	−1/√12	−1/√12
E <sub>y</sub>	.	1/2	.	1/2	−1/2	−1/2
E <sub>y</sub>	2/√12	1/√12	−2/√12	−1/√12	−1/√12	1/√12

As a representative example, we have chosen the combination states with two excited quanta on one of the bendings of *F<sub>2</sub>* symmetry, with orientation *x*, *y*, or *z*, and four quanta on another orientation. Each column of the table corresponds to one of the six possible states  $|\theta_{F_{2x}}^2 \theta_{F_{2y}}^4\rangle = |240\rangle$ ,  $|\theta_{F_{2x}}^2 \theta_{F_{2z}}^4\rangle = |204\rangle$ , ...,  $|\theta_{F_{2y}}^4 \theta_{F_{2z}}^2\rangle = |042\rangle$ . These states (the *B* basis) do not, in general, transform as IRREPs of the *T<sub>d</sub>* group. Each row of the table corresponds to a *C* state, which is a symmetry adapted combination of the *B* states. For each *C* state, we list the IRREP to which it belongs and, in the body of the table, the Clebsh–Gordan transformation coefficients. As discussed in the text, the two *C* states of symmetry *E<sub>y</sub>* can be discarded.

**TABLE II.**  
**Overall Flow of Control in the Build Phase.**

1. Read in the group generators and generate the complete group by closure (Data Structures—System Specification).
2. Read in a minimal subset of atomic coordinates and generate the entire molecule by closure (Data Structures—System Specification).
3. Read in a minimal subset of internal coordinates and generate the entire set by closure (Data Structures—System Specification).
4. Construct the symmetry coordinates by the projection operator method (Theory—Symmetry Coordinates).
5. Read in the order of the basis and the scheme of the resonances (Theory—Choice of the Basis).
6. For each block in the basis:
  - generate a basis of products of harmonic oscillator wave functions in symmetry coordinates (Theory—Choice of the Basis);
  - construct the Clebsh–Gordan coefficients by projection on symmetry adapted components (Theory—Symmetry-Adapted Basis).
7. For each order in the Hamiltonian:
  - identify the distinct Hamiltonian constants in internal coordinates (Theory—Equivalent Hamiltonian Constants);
  - obtain the Hamiltonian in internal coordinates (Theory—Hamiltonian in Internal Coordinates; and Theory—Term Multiplicity);
  - transform the Hamiltonian constants into symmetry coordinates (Theory—Equivalent Hamiltonian Constants);
  - obtain the Hamiltonian in symmetry coordinates (Theory—Term Multiplicity).
8. Save a compact representation of basis, Clebsh–Gordan coefficients and Hamiltonian.

Where appropriate, a reference to a section of the article is given in parentheses.

Because symmetry operators apply to atoms, group operators, internal coordinates, and Hamiltonian constants, we are naturally led to describe these objects in an uniform way. Atoms are first identified by their cartesian coordinates  $(x_1, x_2, x_3)$ , and are then labeled with integers  $a$ . Group operators  $\hat{G}$  are first identified by the matrices  $D^I(\hat{G})$  and  $D(\hat{G})$ , and are then labeled with integers  $g$ . For uniformity with the case of atoms, we regard the matrix components as the identifying coordinates of the operators. Internal coordinates, such as stretchings  $r_{a_1 a_2}$  and bendings  $\theta_{a_1 a_2 a_3}$ , are identified by the label  $a_j$  of the atoms involved and are labeled with integers  $r$ . Therefore the atomic labels  $a_j$  are the identifying coordinates of the internal coordinates  $R(a_1, a_2 \dots)$ . Finally,  $n$ -dimensional Hamiltonian constants  $V_{r_1 r_2 \dots r_n}$  and  $T_{r_1 r_2 \dots r_n}$  are identified by the labels  $r_j$  of the internal coordinates involved. Again, the internal labels  $r_j$  are the identifying coordinates of the Hamiltonian constants.

We have found that this description is very natural to adopt and easy to extend. For the basis vectors, for example, we choose the quantum numbers as identifying coordinates. Symmetry coordinates, symmetry adapted basis vectors, and Hamiltonian constants in symmetry coordinates, which can all be described as linear combinations of other objects, have the combination coefficients as identifying coordinates.

This uniform description suggests a common data structure for all kinds of objects. A single object is represented by its identifying coordinates, which we choose to store into a single vector. Sets of objects can be stored as lists of vectors or, equivalently, as two-dimensional arrays. The choice of a common data structure allows a spectacular reduction of the redundancy and complexity of the program, because many algorithms, including comparison of objects, closure, identification of equivalence classes, projection on symmetry-adapted components, sorting by lexicographic order, orthonormalization, and input and output, become abstract procedures totally independent of the kind of objects that they handle. The action of the group operators on the identifying coordinates of an object does, of course, depend on the kind of object. Matrix by matrix products, or matrix by vector, apply when the objects are group operators identified by their matrices, or atoms identified by their cartesian coordinates. Renaming of the labels applies when the objects are internal coordinates identified by sets of atomic labels, or Hamiltonian constants identified by sets of internal coordinates labels.

From a programmer point of view, the most natural way to represent the intrinsic polymorphism of the group operators is to require objects to provide their own methods for computing the operator action. This object-oriented (OO) paradigm is widely recognized as an effective way of developing robust software. Several modern programming languages, such as C++, Eiffel and Java<sup>30-32</sup> directly support the OO paradigm. For speed of prototyping, we have chosen a C-derived interpreted language, AWK,<sup>33</sup> and augmented it with an OO library. The AWK language supports sparse arrays with fully automatic memory allocation, and has extensive string handling capability. Both facilities have proved crucial for our purposes.

## ORGANIZATION OF THE SOFTWARE PACKAGE

As discussed in the introduction, the package is organized in a "build" phase and in a "deploy" phase. The overall structure of the two phases is summarized in Tables II and III. The "build" phase takes care of molecular symmetry and geometry, construction of the Hamiltonian in a symbolic form, and construction and symmetrization of the basis. This part, which avoids extensive number crunching, has been written in the AWK language<sup>33</sup> with an OO library, as mentioned above. For greater efficiency the code could be rewritten in a compiled OO language like C++.<sup>30</sup> Because the "build" phase is typically repeated only few times, while one verifies that the input description of the problem is correct, the relative inefficiency of this prototype is not of serious concern. In fact, what would have taken days or weeks if done by hand, is now done in a few hours, and without errors.

A compact specification of the problem is sufficient for the "build" phase. To describe the molecule, one has only to supply a minimal subset of the symmetry group, of the atomic coordinates, and of the internal coordinates. The specification of the system is then completed by choosing the highest anharmonic order to be considered and the highest overtone to be calculated. The outputs of the "build" phase are (1) a symbolic representation of the Hamiltonian and of the basis, (2) a machine-oriented encoding of the same information, to be used to drive a spectral fit procedure, and (3) a Fortran<sup>34</sup> source code for the evaluation of the coefficients of the Hamiltonian, to be inserted into the fit procedure.

In the "deploy" phase, the force and kinetic energy constants assume values suitable for the molecule of interest, the matrix elements of the

**TABLE III.**  
**Overall Flow of Control in the Deploy Phase.**

1. Read in force constants and any required molecular parameter (Theory—Equivalent Hamiltonian Constants).
2. Reload the representation of basis, Clebsh–Gordan coefficients and Hamiltonian  $H$ .
3. For each block in the basis (Theory—Choice of the Basis):
  - compute the Hamiltonian matrix between basis vectors in the block by applying intrablock  $H^0$  terms once (Theory—Hamiltonian in a Matrix Form);
  - add second-order perturbation contributions by applying out-of-block  $H^1$  terms twice (Theory—Second-Order Perturbations);
  - use Clebsh–Gordan coefficients to transform the Hamiltonian matrix to symmetry adapted basis (Theory—Symmetry-Adapted Basis);
  - compute vibrational wave numbers by diagonalizing each symmetry adapted block of the transformed Hamiltonian matrix.
4. The fit to the experimental wave numbers, if desired, is performed by altering the force constants, and repeating step 3, as many times as necessary (Conclusions).

Where appropriate, a reference to a section of the article is given in parentheses.

Hamiltonian are computed numerically, the matrix is reduced to a block diagonal form by using the Clebsh–Gordan coefficients, and the vibrational wave numbers are finally determined by diagonalizing the blocks. In most cases, the force constants are considered as parameters to be fitted to the observed wave numbers, and the calculation is to be repeated many times before the fit converges.

Because the “deploy” phase is purely numeric in character, it must be as efficient as possible, and because we wanted to interface it with existing diagonalizing and fitting procedures, we have implemented it in Fortran.<sup>34</sup> The Hamiltonian, basis, and symmetrization routines, which are completely independent of the specific molecular problem, are driven by tables automatically prepared by the “build” phase. The tables describe the Hamiltonian in terms of creation and annihilation operators, the basis in terms of products of harmonic wave functions in symmetry coordinates, and the Clebsh–Gordan coefficients. The only part of the code that depends on the specific molecular problem, namely the transformation of the Hamiltonian constants from internal to symmetry coordinates, is done by a Fortran source mechanically written by the “build” phase.

The fitting code is very flexible and offers several methods: Tsallis–Stariolo<sup>35,36</sup> simulated annealing, which is often able to find a global minimum even in presence of competing local minima; the Nelder–Mead<sup>37</sup> simplex minimization, which does

not require numerical derivatives and is, therefore, quite fast at the initial stages of a refinement; the Levenberg–Marquardt<sup>37</sup> algorithm, for the most simple cases; and singular value<sup>37</sup> decomposition, for ill-conditioned problems with undetermined or highly correlated force constants. In such cases, the method will converge in a sensible way towards a minimum, while producing an analysis of the cross-correlation.

When information on the potential energy surface is available from *ab initio* calculations, previous fits on different isotopomers, or other sources, it is possible to add it into the fit, and obtain a well-determined model still in agreement with the experiments. To achieve this,<sup>4</sup> we have adopted the method of predicate observations,<sup>3,38</sup> which can be described as a Bayesian approach,<sup>39</sup> in which experimental and *a priori* information (known force constants) are combined.

## Conclusions

We have presented the “build & deploy” (B&D) software package for the calculation of anharmonic vibrational spectra of nonlinear molecules. The package was assembled in several months of work on a IBM-RS/6000 machine, mostly using public domain software tools, and has been ported without unexpected problems to other Unix platforms. We feel that the effort spent in constructing the pack-

age, even in its first application,<sup>4</sup> pays back for the time required by a more conventional programming strategy. Extension to higher overtones, for instance, now requires a few minutes, while a new molecule problem could be solved within hours. The current release of the package, which is still somehow experimental, is available by contacting the authors. The User's Manual, example input and output files are available via the [www](http://www.5).<sup>5</sup>

## Acknowledgments

E. V. wishes to thank the EC for further support (FMRX-CT96-0088, TMR network).

## Appendix A: Example Hamiltonian

As a typical example of the Hamiltonians that can be handled by the "B&D" program, we report in this appendix the quartic potential operator  $V_{rr\theta\theta}$  involving two stretching coordinates  $r$  and two bending coordinates  $\theta$  for the  $\text{CH}_4$  molecule. There are four  $\text{C—H}_i$  stretching coordinates  $r_i$ , with  $1 \leq i \leq 4$ , and six  $\text{H}_i\text{—C—H}_j$  bending coordinates  $\theta_{ij}$ , with  $1 \leq i < j \leq 4$ . The algorithm described in the text and illustrated in Figure 1 yields 15 equivalence classes for the force constants  $F_{r_i r_j \theta_{kl} \theta_{mn}}$  involving two stretchings and two bendings. The  $V_{rr\theta\theta}$  operator [eq. (6)] expressed in internal coordinates is:

$$\begin{aligned} V_{rr\theta\theta} &= 6F_{r_1 r_1 \theta_{12} \theta_{12}} (r_1 r_1 \theta_{12} \theta_{12} + r_1 r_1 \theta_{13} \theta_{13} + r_1 r_1 \theta_{14} \theta_{14} \\ &\quad + r_2 r_2 \theta_{12} \theta_{12} + r_2 r_2 \theta_{23} \theta_{23} + r_2 r_2 \theta_{24} \theta_{24} \\ &\quad + r_3 r_3 \theta_{13} \theta_{13} + r_3 r_3 \theta_{23} \theta_{23} + r_3 r_3 \theta_{34} \theta_{34} \\ &\quad + r_4 r_4 \theta_{14} \theta_{14} + r_4 r_4 \theta_{24} \theta_{24} + r_4 r_4 \theta_{34} \theta_{34}) \\ &+ 12F_{r_1 r_1 \theta_{12} \theta_{13}} (r_1 r_1 \theta_{12} \theta_{13} + r_1 r_1 \theta_{12} \theta_{14} + r_1 r_1 \theta_{13} \theta_{14} \\ &\quad + r_2 r_2 \theta_{12} \theta_{23} + r_2 r_2 \theta_{12} \theta_{24} + r_2 r_2 \theta_{23} \theta_{24} \\ &\quad + r_3 r_3 \theta_{13} \theta_{23} + r_3 r_3 \theta_{13} \theta_{34} + r_3 r_3 \theta_{23} \theta_{34} \\ &\quad + r_4 r_4 \theta_{14} \theta_{24} + r_4 r_4 \theta_{14} \theta_{34} + r_4 r_4 \theta_{24} \theta_{34}) \\ &+ 12F_{r_1 r_1 \theta_{12} \theta_{23}} (r_1 r_1 \theta_{12} \theta_{23} + r_1 r_1 \theta_{12} \theta_{24} + r_1 r_1 \theta_{13} \theta_{23} \\ &\quad + r_1 r_1 \theta_{13} \theta_{34} + r_1 r_1 \theta_{14} \theta_{24} + r_1 r_1 \theta_{14} \theta_{34} \\ &\quad + r_2 r_2 \theta_{12} \theta_{13} + r_2 r_2 \theta_{12} \theta_{14} + r_2 r_2 \theta_{13} \theta_{23} \\ &\quad + r_2 r_2 \theta_{14} \theta_{24} + r_2 r_2 \theta_{23} \theta_{34} + r_2 r_2 \theta_{24} \theta_{34} \\ &\quad + r_3 r_3 \theta_{12} \theta_{13} + r_3 r_3 \theta_{12} \theta_{23} + r_3 r_3 \theta_{13} \theta_{14} \\ &\quad + r_3 r_3 \theta_{14} \theta_{34} + r_3 r_3 \theta_{23} \theta_{24} + r_3 r_3 \theta_{24} \theta_{34} \\ &\quad + r_4 r_4 \theta_{12} \theta_{14} + r_4 r_4 \theta_{12} \theta_{24} + r_4 r_4 \theta_{13} \theta_{14} \\ &\quad + r_4 r_4 \theta_{13} \theta_{34} + r_4 r_4 \theta_{23} \theta_{24} + r_4 r_4 \theta_{23} \theta_{34}) \end{aligned}$$

$$\begin{aligned} &+ 12F_{r_1 r_1 \theta_{12} \theta_{34}} (r_1 r_1 \theta_{12} \theta_{34} + r_1 r_1 \theta_{13} \theta_{24} + r_1 r_1 \theta_{14} \theta_{23} \\ &\quad + r_2 r_2 \theta_{12} \theta_{34} + r_2 r_2 \theta_{13} \theta_{24} + r_2 r_2 \theta_{14} \theta_{23} \\ &\quad + r_3 r_3 \theta_{12} \theta_{34} + r_3 r_3 \theta_{13} \theta_{24} + r_3 r_3 \theta_{14} \theta_{23} \\ &\quad + r_4 r_4 \theta_{12} \theta_{34} + r_4 r_4 \theta_{13} \theta_{24} + r_4 r_4 \theta_{14} \theta_{23}) \\ &+ 6F_{r_1 r_1 \theta_{23} \theta_{23}} (r_1 r_1 \theta_{23} \theta_{23} + r_1 r_1 \theta_{24} \theta_{24} + r_1 r_1 \theta_{34} \theta_{34} \\ &\quad + r_2 r_2 \theta_{13} \theta_{13} + r_2 r_2 \theta_{14} \theta_{14} + r_2 r_2 \theta_{34} \theta_{34} \\ &\quad + r_3 r_3 \theta_{12} \theta_{12} + r_3 r_3 \theta_{14} \theta_{14} + r_3 r_3 \theta_{24} \theta_{24} \\ &\quad + r_4 r_4 \theta_{12} \theta_{12} + r_4 r_4 \theta_{13} \theta_{13} + r_4 r_4 \theta_{23} \theta_{23}) \\ &+ 12F_{r_1 r_1 \theta_{23} \theta_{24}} (r_1 r_1 \theta_{23} \theta_{24} + r_1 r_1 \theta_{23} \theta_{34} + r_1 r_1 \theta_{24} \theta_{34} \\ &\quad + r_2 r_2 \theta_{13} \theta_{14} + r_2 r_2 \theta_{13} \theta_{34} + r_2 r_2 \theta_{14} \theta_{34} \\ &\quad + r_3 r_3 \theta_{12} \theta_{14} + r_3 r_3 \theta_{12} \theta_{24} + r_3 r_3 \theta_{14} \theta_{24} \\ &\quad + r_4 r_4 \theta_{12} \theta_{13} + r_4 r_4 \theta_{12} \theta_{23} + r_4 r_4 \theta_{13} \theta_{23}) \\ &+ 12F_{r_1 r_2 \theta_{12} \theta_{12}} (r_1 r_2 \theta_{12} \theta_{12} + r_1 r_3 \theta_{13} \theta_{13} + r_1 r_4 \theta_{14} \theta_{14} \\ &\quad + r_2 r_3 \theta_{23} \theta_{23} + r_2 r_4 \theta_{24} \theta_{24} + r_3 r_4 \theta_{34} \theta_{34}) \\ &+ 24F_{r_1 r_2 \theta_{12} \theta_{13}} (r_1 r_2 \theta_{12} \theta_{13} + r_1 r_2 \theta_{12} \theta_{14} + r_1 r_2 \theta_{12} \theta_{23} \\ &\quad + r_1 r_2 \theta_{12} \theta_{24} + r_1 r_3 \theta_{13} \theta_{13} + r_1 r_3 \theta_{13} \theta_{14} \\ &\quad + r_1 r_3 \theta_{13} \theta_{23} + r_1 r_3 \theta_{13} \theta_{34} + r_1 r_4 \theta_{12} \theta_{14} \\ &\quad + r_1 r_4 \theta_{13} \theta_{14} + r_1 r_4 \theta_{14} \theta_{24} + r_1 r_4 \theta_{14} \theta_{34} \\ &\quad + r_2 r_3 \theta_{12} \theta_{23} + r_2 r_3 \theta_{13} \theta_{23} + r_2 r_3 \theta_{23} \theta_{24} \\ &\quad + r_2 r_3 \theta_{23} \theta_{34} + r_2 r_4 \theta_{12} \theta_{24} + r_2 r_4 \theta_{14} \theta_{24} \\ &\quad + r_2 r_4 \theta_{23} \theta_{24} + r_2 r_4 \theta_{24} \theta_{34} + r_3 r_4 \theta_{13} \theta_{34} \\ &\quad + r_3 r_4 \theta_{14} \theta_{34} + r_3 r_4 \theta_{23} \theta_{34} + r_3 r_4 \theta_{24} \theta_{34}) \\ &+ 24F_{r_1 r_2 \theta_{12} \theta_{34}} (r_1 r_2 \theta_{12} \theta_{34} + r_1 r_3 \theta_{13} \theta_{24} + r_1 r_4 \theta_{14} \theta_{23} \\ &\quad + r_2 r_3 \theta_{14} \theta_{23} + r_2 r_4 \theta_{13} \theta_{24} + r_3 r_4 \theta_{12} \theta_{34}) \\ &+ 12F_{r_1 r_2 \theta_{13} \theta_{13}} (r_1 r_2 \theta_{13} \theta_{13} + r_1 r_2 \theta_{14} \theta_{14} + r_1 r_2 \theta_{23} \theta_{23} \\ &\quad + r_1 r_2 \theta_{24} \theta_{24} + r_1 r_3 \theta_{12} \theta_{12} + r_1 r_3 \theta_{14} \theta_{14} \\ &\quad + r_1 r_3 \theta_{23} \theta_{23} + r_1 r_3 \theta_{34} \theta_{34} + r_1 r_4 \theta_{12} \theta_{12} \\ &\quad + r_1 r_4 \theta_{13} \theta_{13} + r_1 r_4 \theta_{24} \theta_{24} + r_1 r_4 \theta_{34} \theta_{34} \\ &\quad + r_2 r_3 \theta_{12} \theta_{12} + r_2 r_3 \theta_{13} \theta_{13} + r_2 r_3 \theta_{24} \theta_{24} \\ &\quad + r_2 r_3 \theta_{34} \theta_{34} + r_2 r_4 \theta_{12} \theta_{12} + r_2 r_4 \theta_{14} \theta_{14} \\ &\quad + r_2 r_4 \theta_{23} \theta_{23} + r_2 r_4 \theta_{34} \theta_{34} + r_3 r_4 \theta_{13} \theta_{13} \\ &\quad + r_3 r_4 \theta_{14} \theta_{14} + r_3 r_4 \theta_{23} \theta_{23} + r_3 r_4 \theta_{24} \theta_{24}) \\ &+ 24F_{r_1 r_2 \theta_{13} \theta_{14}} (r_1 r_2 \theta_{13} \theta_{14} + r_1 r_2 \theta_{23} \theta_{24} + r_1 r_3 \theta_{12} \theta_{14} \\ &\quad + r_1 r_3 \theta_{23} \theta_{34} + r_1 r_4 \theta_{12} \theta_{13} + r_1 r_4 \theta_{24} \theta_{34} \\ &\quad + r_2 r_3 \theta_{12} \theta_{24} + r_2 r_3 \theta_{13} \theta_{34} + r_2 r_4 \theta_{12} \theta_{23} \\ &\quad + r_2 r_4 \theta_{14} \theta_{34} + r_3 r_4 \theta_{13} \theta_{23} + r_3 r_4 \theta_{14} \theta_{24}) \\ &+ 24F_{r_1 r_2 \theta_{13} \theta_{23}} (r_1 r_2 \theta_{13} \theta_{23} + r_1 r_2 \theta_{14} \theta_{24} + r_1 r_3 \theta_{12} \theta_{23} \\ &\quad + r_1 r_3 \theta_{14} \theta_{34} + r_1 r_4 \theta_{12} \theta_{24} + r_1 r_4 \theta_{13} \theta_{34} \\ &\quad + r_2 r_3 \theta_{12} \theta_{13} + r_2 r_3 \theta_{24} \theta_{34} + r_2 r_4 \theta_{12} \theta_{14} \\ &\quad + r_2 r_4 \theta_{23} \theta_{34} + r_3 r_4 \theta_{13} \theta_{14} + r_3 r_4 \theta_{23} \theta_{24}) \\ &+ 24F_{r_1 r_2 \theta_{13} \theta_{24}} (r_1 r_2 \theta_{13} \theta_{24} + r_1 r_2 \theta_{14} \theta_{23} + r_1 r_3 \theta_{12} \theta_{34} \\ &\quad + r_1 r_3 \theta_{14} \theta_{23} + r_1 r_4 \theta_{12} \theta_{34} + r_1 r_4 \theta_{13} \theta_{24}) \end{aligned}$$

$$\begin{aligned}
& + r_2 r_3 \theta_{12} \theta_{34} + r_2 r_3 \theta_{13} \theta_{24} + r_2 r_4 \theta_{12} \theta_{34} \\
& + r_2 r_4 \theta_{14} \theta_{23} + r_3 r_4 \theta_{13} \theta_{24} + r_3 r_4 \theta_{14} \theta_{23}) \\
& + 24F_{r_1 r_2 \theta_{13} \theta_{34}} (r_1 r_2 \theta_{13} \theta_{34} + r_1 r_2 \theta_{14} \theta_{34} + r_1 r_2 \theta_{23} \theta_{34} \\
& + r_1 r_2 \theta_{24} \theta_{34} + r_1 r_3 \theta_{12} \theta_{24} + r_1 r_3 \theta_{14} \theta_{24} \\
& + r_1 r_3 \theta_{23} \theta_{24} + r_1 r_3 \theta_{24} \theta_{34} + r_1 r_4 \theta_{12} \theta_{23} \\
& + r_1 r_4 \theta_{13} \theta_{23} + r_1 r_4 \theta_{23} \theta_{24} + r_1 r_4 \theta_{23} \theta_{34} \\
& + r_2 r_3 \theta_{12} \theta_{14} + r_2 r_3 \theta_{13} \theta_{14} + r_2 r_3 \theta_{14} \theta_{24} \\
& + r_2 r_3 \theta_{14} \theta_{34} + r_2 r_4 \theta_{12} \theta_{13} + r_2 r_4 \theta_{13} \theta_{14} \\
& + r_2 r_4 \theta_{13} \theta_{23} + r_2 r_4 \theta_{13} \theta_{34} + r_3 r_4 \theta_{12} \theta_{13} \\
& + r_3 r_4 \theta_{12} \theta_{14} + r_3 r_4 \theta_{12} \theta_{23} + r_3 r_4 \theta_{12} \theta_{24}) \\
& + 12F_{r_1 r_2 \theta_{34} \theta_{34}} (r_1 r_2 \theta_{34} \theta_{34} + r_1 r_3 \theta_{24} \theta_{24} + r_1 r_4 \theta_{23} \theta_{23} \\
& + r_2 r_3 \theta_{14} \theta_{14} + r_2 r_4 \theta_{13} \theta_{13} + r_3 r_4 \theta_{12} \theta_{12})
\end{aligned} \quad (A1)$$

As discussed in the text, the first constant in each equivalence class is chosen to represent all the others. The numerical coefficients represent the term multiplicity as described in the Theory—The Hamiltonian section. The multiplicity associated to  $F_{r_i r_j \theta_{kl} \theta_{mn}}$  is  $4! = 24$  if  $r_i \neq r_j$  and  $\theta_{kl} \neq \theta_{mn}$ . The multiplicity must be divided by  $2! = 2$  if  $r_i = r_j$  or  $\theta_{kl} = \theta_{mn}$ , and by  $2!2! = 4$  if  $r_i = r_j$  and  $\theta_{kl} = \theta_{mn}$ .

We choose the matrices  $D_{ij}^l(\hat{G})$  for the irreducible representation of the  $T_d$  symmetry group so that they reproduce the standard symmetry coordinates for  $\text{CH}_4$ .<sup>40</sup> Due to the redundancy associated with the bendings,<sup>3,20</sup> there are only nine independent symmetry coordinates. Following the spectroscopic notation for the normal modes  $\nu_1, \nu_2, \nu_3, \nu_4$ , these coordinates are labeled  $S_1$  ( $A_1$  stretching),  $S_{2a}, S_{2b}$  ( $E$  bending),  $S_{3x}, S_{3y}, S_{3z}$  ( $F_2$  stretching), and  $S_{4x}, S_{4y}, S_{4z}$  ( $F_2$  bending). The symmetry coordinates as linear combinations of internal coordinates [eq. (9)] and the equivalence classes for the force constants yield the  $V_{rr\theta\theta}$  Hamiltonian in symmetry coordinates, using eqs. (15), (14), and (12) in sequence. The  $V_{rr\theta\theta}$  operator thus becomes:

$$\begin{aligned}
V_{rr\theta\theta} = & 6F_{112a2a}(S_1 S_1 S_{2a} S_{2a} + S_1 S_1 S_{2b} S_{2b}) \\
& + 6F_{114x4x}(S_1 S_1 S_{4x} S_{4x} + S_1 S_1 S_{4y} S_{4y} \\
& + S_1 S_1 S_{4z} S_{4z}) \\
& + 24F_{12a3x4x}(S_1 S_{2a} S_{3x} S_{4x} + \sqrt{3} S_1 S_{2b} S_{3x} S_{4x} \\
& + S_1 S_{2a} S_{3y} S_{4y} + \sqrt{3} S_1 S_{2b} S_{3y} S_{4y} \\
& - 2S_1 S_{2a} S_{3z} S_{4z}) \\
& + 24F_{13x4y4z}(S_1 S_{3x} S_{4y} S_{4z} + S_1 S_{3y} S_{4x} S_{4z} \\
& + S_1 S_{3z} S_{4x} S_{4y}) \\
& + 6F_{2a2a3x3x}(S_{2a} S_{2a} S_{3x} S_{3x} + S_{2a} S_{2a} S_{3y} S_{3y}) \\
& + 12F_{2a2b3x3x}(S_{2a} S_{2b} S_{3x} S_{3x} - S_{2a} S_{2b} S_{3y} S_{3y}) \\
& + 6F_{2b2b3x3x}(S_{2b} S_{2b} S_{3x} S_{3x} + S_{2b} S_{2b} S_{3y} S_{3y})
\end{aligned}$$

$$\begin{aligned}
& + 6F_{3x3x4x4x}(S_{3x} S_{3x} S_{4x} S_{4x} + S_{3y} S_{3y} S_{4y} S_{4y} \\
& + S_{3z} S_{3z} S_{4z} S_{4z}) \\
& + 6F_{3x3x4y4y}(S_{3x} S_{3x} S_{4y} S_{4y} + S_{3x} S_{3x} S_{4z} S_{4z} \\
& + S_{3y} S_{3y} S_{4x} S_{4x} + S_{3y} S_{3y} S_{4z} S_{4z} \\
& + S_{3z} S_{3z} S_{4x} S_{4x} + S_{3z} S_{3z} S_{4y} S_{4y}) \\
& + 12F_{2a3x3y4z}(2S_{2a} S_{3x} S_{3y} S_{4z} - S_{2a} S_{3x} S_{3z} S_{4y} \\
& - \sqrt{3} S_{2b} S_{3x} S_{3z} S_{4y} - S_{2a} S_{3y} S_{3z} S_{4x} \\
& + \sqrt{3} S_{2b} S_{3y} S_{3z} S_{4x}) \\
& + 24F_{3x3y4x4y}(S_{3x} S_{3y} S_{4x} S_{4y} + S_{3x} S_{3z} S_{4x} S_{4z} \\
& + S_{3y} S_{3z} S_{4y} S_{4z}) \\
& + 6F_{2a2a3z3z}(S_{2a} S_{2a} S_{3z} S_{3z}) \\
& + 6F_{2b2b3z3z}(S_{2b} S_{2b} S_{3z} S_{3z})
\end{aligned} \quad (A2)$$

The  $V_{rr\theta\theta}$  potential contains 13 force constants expressed in symmetry coordinates, but only 10 of them are actually independent,<sup>20</sup> due to the occurrence of linear dependencies (as discussed in the text). The dependencies found by the program are:

$$\begin{aligned}
F_{2b2b3x3x} &= (\sqrt{3}F_{2a2a3x3x} - 2F_{2a2b3x3x})/\sqrt{3} \\
F_{2a2a3z3z} &= (F_{2a2a3x3x} - \sqrt{3}F_{2a2b3x3x}) \\
F_{2b2b3z3z} &= (\sqrt{3}F_{2a2a3x3x} + F_{2a2b3x3x})/\sqrt{3}
\end{aligned} \quad (A3)$$

It should be noticed that the Hamiltonian has been determined with purely numerical methods, by manipulating the Jacobian matrix as described in the text. However, it is possible to obtain exactly the same results analytically by replacing the symmetry coordinates in the Hamiltonian expressed in internal coordinates, as we have verified at lower orders for  $\text{CH}_4$ , and at the fourth order for smaller molecules.

## References

1. Lukka, T.; Kauppi, E.; Halonen, L. *J Chem Phys* 1995, 102, 5200.
2. Kauppi, E.; Halonen, L. *J Chem Phys* 1995, 103, 6861.
3. Halonen, L. *J Chem Phys* 1997, 106, 831.
4. Venuti, E.; Halonen, L.; Della Valle, R. G. *J Chem Phys* 1999, 110, 7339.
5. Della Valle, R. G.; Halonen, L.; Venuti, E., B&D User's Manual, <http://www.chim.unifi.it:8080/~valle/>
6. Wilson, E. B.; Decius, J. C.; Cross, P. C. *Molecular Vibrations*; McGraw-Hill: New York, 1955.
7. Califano, S. *Vibrational States*; Wiley: London, 1976.
8. Hoy, A. R.; Mills, I. M.; Strey, G. *Mol Phys* 1972, 24, 1265.
9. Pickett, H. M. *J Chem Phys* 1972, 56, 1715.
10. Halonen, L.; Carrington Jr., T. *J Chem Phys* 1987, 88, 4171.
11. Messiah, A. *Quantum Mechanics*; Wiley: New York, 1961.
12. Bacic, Z.; Light, J. C. *Annu Rev Phys Chem* 1989, 40, 469.

13. Tennyson, J.; Henderson, J. R.; Fulton, N. G. *Computer Phys Commun* 1995, 86, 175.
14. Graver, J. E.; Watkins, M. E. *Combinatorics with Emphasis on the Theory of Graphs*; Springer Verlag: Berlin, 1977.
15. Allburn, E. Dr. *Dobb's J* 1991, 172, 88.
16. Knuth, D. E. *The Art of Computer Programming, Vol. 1, Fundamental Algorithms*; Addison-Wesley: Reading, MA, 1981.
17. Atkins, P. W.; Friedman, R. S. *Molecular Quantum Mechanics*; Oxford University: Oxford, 1997.
18. Cornwell, J. F. *Group Theory in Physics*; Academic Press: London, 1984.
19. Apostol, T. M. *Calculus*; Wiley: New York, 1967.
20. Raynes, W. T.; Lazzaretti, P.; Zanasi, R.; Sadley, A. J.; Fowler, P. W. *Mol Phys* 1987, 60, 509.
21. Watson, J. K. G. *J Mol Spectrosc* 1972, 41, 229.
22. Henry, L.; Amat, G. *J Mol Spectrosc* 1960, 5, 319; *ibid* 1965, 15, 168.
23. Knuth, D. E. *The Art of Computer Programming, Vol. 3, Sorting and Searching*; Addison-Wesley: Reading, MA, 1973.
24. Landau, L. D.; Lifchitz, E. *Quantum Mechanics*; Pergamon: Oxford, 1958.
25. Mills, I. M.; Mompean, F. J. *Chem Phys Lett* 1986, 124, 425.
26. Della Valle, R. G.; Procacci, P. *Phys Rev B* 1992, 46, 6141.
27. Procacci, P.; Signorini, G. F.; Della Valle, R. G. *Phys Rev B* 1993, 47, 11124.
28. Sibert, E. L. *Comp Phys Commun* 1988, 51, 149.
29. McCoy, A. B.; Sibert, E. L. In: Wyatt, R. E.; Zhang, J. Z. H., eds. *Dynamics of Molecules and Chemical Reactions*; Marcel Dekker Inc.: New York, 1996.
30. Stroustrup, B. *The C++ Programming Language*; Addison-Wesley: Reading, MA, 1991.
31. Meyer, B. *Eiffel: The Language*; Prentice-Hall: New York, 1992.
32. Arnold K.; Gosling, J. *The Java Programming Language*; Addison-Wesley: Reading, MA, 1996.
33. Aho, A. V.; Kernighan, B. W.; Weinberger, P. J. *The AWK Programming Language*; Addison-Wesley: Reading, MA, 1988.
34. Ellis, T. M. R. *Fortran 77 Programming: with an Introduction to Fortran 90 Standard*; Addison-Wesley: Reading, MA, 1990.
35. Tsallis, C.; Stariolo, D. A. *Physica A* 1996, 233, 395.
36. Hansmann, U. H. E. *Physica A* 1997, 242, 250.
37. Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; Vetterling, W. T. *Numerical Recipes*; Cambridge University Press: Cambridge, 1986.
38. Bartell, L. S.; Romenesko, D. J.; Wong, T. C. In: Sim, G. A.; Sutton, L. E., eds. *Specialist Periodical Report, Molecular Structure by Diffraction Methods*; The Chemical Society: London, 1975, Vol. 3.
39. Sivia, D. S. *Data Analysis, A Bayesian Tutorial*; Clarendon Press: Oxford, 1996.
40. Gray, D. S.; Robiette, A. G. *Mol Phys* 1979, 37, 1901.